DOI:
10.1016/j.csda.2016.05.023

Document Version
Peer reviewed version

Factorial and response surface designs robust to missing observations

Marcelo A. da Silva

Programa Interinstitutional de Pós-Graduação em Estatística, USP/UFSCar, São Carlos, Brazil

Steven G. Gilmour

Department of Mathematics, King's College London, UK

Luzia A. Trinca*

Departamento de Bioestatística, IB, Universidade Estadual Paulista, Botucatu, Brazil

Abstract

Compound optimum design criteria which allow pure error degrees of freedom may produce designs that break down when even a single run is missing, if the number of experimental units is small. The inclusion, in the compound criteria, of a measure of leverage uniformity is proposed in order to produce designs that are more robust to missing observations. By appropriately choosing the weights of each part of the criterion, robust designs are obtained that are also highly efficient in terms of other properties. Applications to various experimental setups show the advantages of the new methods.

Keywords: Compound criteria, Cook's distance, leverage, optimum design, pure error

*Corresponding Author

Email addresses: silva.marcelo@usp.br (Marcelo A. da Silva), steven.gilmour@kcl.ac.uk (Steven G. Gilmour), ltrinca@ibb.unesp.br (Luzia A. Trinca)

1. Introduction

Processes, products and methods in many areas are discovered and improved by performing controlled experiments in which the levels of several continuous inputs, experimental factors, are manipulated and at least one outcome is measured. Empirical models, such as low order polynomials, relating the response to the factor levels have been extremely useful for interpreting the data from such experiments. Such models and methods are part of the large area of Response Surface Methodology. Designs for experiments in this set up are known as Response Surface (RS) designs.

It has long been recognized that the experimental design should have several good properties. In the context of RS, Box and Draper (1975) started a list that was subsequently enlarged (Box and Draper, 1987, 2007) to 14 desired properties, some of them conflicting, indicating that in practice it is wise and necessary to compromise in order to choose a good design.

On the other hand, optimum design methodologies have concentrated on variance-based criteria such as D-, A- and I-optimality, the so called alphabetical optimality, see Atkinson et al. (2007) for an account of design criterion definitions. The use of a single optimality criterion may lead to designs that lack practical appeal. Gilmour and Trinca (2012) re-defined the alphabetical optimality criteria such that their properties are valid under inferences based on the randomization process only. They proposed adjustments to the traditional criteria allowing for pure error degrees of freedom in order to appropriately estimate random variation, the so called DP and AP criteria for instance. However, as recognised by the authors, these criteria may produce extreme designs with no spare degrees of freedom for inclusion of additional model terms. They further proposed compound criteria that aggregate into a single function the properties reflecting four experimental objectives, including a simple component, based on degree of freedom efficiency (Daniel, 1976) to drive the design to allow some lack of fit degrees of freedom as well. The use of compound criteria as well as procedures for multiple objectives (Lu, Anderson-Cook, and Robinson, 2011) has the power to produce designs that are very statistically efficient and useful for experimenters.

Concerning the extreme designs produced by using a single property, e.g. DP, it was pointed out by Ridout (2012) that small designs would break down in case of even one missing observation from some treatment units. Robustness to missing observations is closely related to insensitivity to wild observations, a desired design property highlighted by Box and Draper. A

design is said to be *robust* to missing observations if the model parameters are still estimable, without too much loss of precision, when observations from some experimental units are not available. Just as there are different design optimality criteria for estimation and for inference, there are different criteria for robustness. Surrogate measures related to the so called *leverages*, associated with a regression model, have been used to compare designs in this sense, as well as measures related to precision.

For example, Box and Draper (1975) studied the connections of the sum of squares of leverages and other design measures and found the best replication of center points and axial point values in central composite designs (CCD). Herzberg and Andrews (1976) and Andrews and Herzberg (1979) noted that such a measure does not discriminate well between designs and proposed extended measures incorporating some probability for the event of a missing observation. Akhtar and Prescott (1986) developed an efficiency measure relating the D criterion and the leverages and compared several CCDs, while Ahmad and Gilmour (2010) studied efficiency loss with respect to several optimality criteria due to missing data from different types of points in subset designs (Gilmour, 2006) and Ahmad et al. (2012) did the same for augmented pairs designs. Related investigations were also presented by Ghosh (1982a,b) who studied robustness of certain designs under sets of s missing runs and found the maximum s value for given designs. Adding to these works, Ghosh (1983, 1989) proposed measures to study influence on estimation and prediction of observations. To the best of our knowledge, a property related to robustness to missing data has not yet been incorporated in a criterion function in order to algorithmically construct an efficient RS design robust to missing data.

In this paper we incorporate a measure for the contribution of leverages, related to Cook's distance, in a compound design criterion in order to prevent the optimal design from being too sensitive to some observations or to breakdown in case of missing data. We show through several examples that such a property is particularly important in the case of limited experimental resources. In Section 2, a criterion for assessing design robustness is developed and in Section 3 a brief description of the algorithm is presented. The proposed criterion is shown to work well in several illustrative experiments in Section 4. Some final comments are made in Section 5.

2. Efficient and Robust Designs

Consider a completely randomized design in which there are t treatments, the distinct combinations of levels of q quantitative factors, to be allocated to n experimental units ($t < n$), treatment r being replicated n_r times ($n_r \in \mathbb{N}$, $\sum_{r=1}^{t} n_r = n$). The underlying model for the continuous random response variable Y is

$$y_{rj} = \mu_r + \varepsilon_{rj} \quad r = 1,\ 2,\ \ldots,\ t \quad j = 1,\ 2,\ \ldots,\ n_r, \tag{1}$$

where, in matrix notation, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbf{V}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. Once data are collected the fitting of this model allows $d = n-t$ pure error degrees of freedom to estimate σ^2 unbiasedly. As argued in Box and Draper (2007), in RS experiments we want to simplify the model and add interpretability by approximating

$$\mu_r \approx \mathbf{f}(\mathbf{x}_r)'\boldsymbol{\beta} \quad r = 1,\ 2,\ \ldots,\ t, \tag{2}$$

where $\mathbf{x}_r$ is the vector of levels of the q factors defining treatment r (the design experimental points), $\mathbf{f}$ is the function that expands the levels according to the desired approximating function, usually a low order polynomial, and $\boldsymbol{\beta}$ is the p-dimensional vector of parameters with its first element being the intercept denoted by β_0. In matrix notation, let $\mathbf{X}_p = [\mathbf{1}|\mathbf{X}]$ be the $n \times p$ model matrix for equation (2) where each row of $\mathbf{X}_p$ corresponding to treatment r is $\mathbf{f}(\mathbf{x}_r)'$, $\mathbf{1}$ is the n dimensional column vector with all elements equal to 1 and $\mathbf{X}$ is a $n \times (p-1)$ matrix.

For the DP design criterion (Gilmour and Trinca, 2012) we should minimise $(F_{p,d;1-\alpha})^p / |\mathbf{X}_p'\mathbf{X}_p|$, where $F_{p,d;1-\alpha}$ is the $1-\alpha$ quantile of the F distribution with p numerator and d denominator degrees of freedom and $1-\alpha$ is the confidence level of the confidence region for the p-parameter vector $\boldsymbol{\beta}$. Other alphabetical optimalities can be defined similarly. Using the $(DP)_S$ criterion, for the case of interest in a subset of p_2 ($p_2 < p$) parameters we should minimise $(F_{p_2,d;1-\alpha})^{p_2} |(\mathbf{M}^{-1})_{22}|$, where $\mathbf{M} = \mathbf{X}_p'\mathbf{X}_p$ and $(\mathbf{M}^{-1})_{22}$ is the portion of $\mathbf{M}^{-1}$ referring to the subset of p_2 parameters of interest. See Atkinson et al. (2007) for details on D_S and other useful design criteria. If we use $p_2 = p-1$ which drops only the intercept (β_0) from the set of parameters of interest, minimising $|(\mathbf{M}^{-1})_{22}|$ is equivalent to maximising $|\mathbf{X}'\mathbf{Q}\mathbf{X}|$, where $\mathbf{Q} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$ and $\mathbf{I}$ is the $n \times n$ identity matrix. Focusing on four design objectives, each with a priority weight κ_l ($\sum_{l=1}^{4} \kappa_l = 1$), representing

1. global F test for treatment effects in $\boldsymbol{\beta}$, with significance level α_1;

2. partial confidence intervals for each regression parameter each with confidence level of $1 - \alpha_2$;
3. point estimation of each regression parameter; and
4. lack of fit degrees of freedom,

Gilmour and Trinca (2012) formulated the compound criterion

$$\frac{|\mathbf{X}'\mathbf{Q}\mathbf{X}|^{\frac{\kappa_1}{p-1}}(n-d)^{\kappa_4}}{(F_{p-1,d;1-\alpha_1})^{\kappa_1}(F_{1,d;1-\alpha_2})^{\kappa_2}[tr\{\mathbf{W}(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\}]^{\kappa_2+\kappa_3}}, \tag{3}$$

where $\mathbf{W}$ is a diagonal matrix of weights for a weighted-A criterion and α_1 and α_2 are the significance levels used in objectives 1 and 2, respectively.

As discussed in Ridout (2012), for small n, considering only the first objective may result in designs that break down in the case of a single missing observation. For design breakdown we use the same meaning as Ghosh (1982a,b), that is, the rank of the $\mathbf{M}$ matrix relative to the reduced design is less than p, if the data from some experimental unit is lost, and thus it is not possible to estimate all the elements of $\boldsymbol{\beta}$. One step in the direction of constructing efficient designs with respect to several properties and simultaneously safeguarding against the related missing observation problems is incorporating in the compound criterion some measure to guide the design search in this respect. Missing observations are a fairly common problem in response surface studies, since many of the combinations of factor levels will never have been studied before and might lead to no response being possible. We will see an example of this in practice in Section 4.

Ghosh (1982a) defined a design to be robust against the unavailability of s observations if the design does not break down after omitting any set of s runs. A complete investigation of robustness would require evaluation of the design under each possible set of s missing data points, which can become very computationally intensive. In practice, if the experimenter faces a situation of high risk of having more than one or two missing observations, he or she should be prepared to start with a reasonably large n, in which case, the sensitivity of the design to unavailable data should be low. Our development, based on well known results in linear models, leads to the use of a surrogate measure of sensitivity that will prevent the design from breaking down when *a single* observation goes missing.

From least squares theory, e.g. Hoaglin and Welsh (1978), if observation i, is removed from the data, the covariance matrix of $\widehat{\boldsymbol{\beta}}_{(-i)}$, the new estimator

of $\boldsymbol{\beta}$, except for the constant σ^2, is

$$\mathbf{M}_{(-i)}^{-1} = (\mathbf{X}'_{p(-i)}\mathbf{X}_{p(-i)})^{-1} = \mathbf{M}^{-1} + \frac{\mathbf{M}^{-1}\mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)'\mathbf{M}^{-1}}{1-h_i}, \tag{4}$$

where h_i, called *leverage*, is the i^{th} element of the diagonal of the projection or *hat* matrix given by $\mathbf{H} = \mathbf{X}_p\mathbf{M}^{-1}\mathbf{X}'_p$.

It is well known that $\text{rank}(\mathbf{H}) = \text{trace}(\mathbf{H}) = p$ and $\frac{1}{n} \leq h_i \leq 1$. The ideal h_i value is p/n ($i = 1,\ 2,\ \ldots,\ n$) in which case the contribution from each observation to estimate its response is the same for all points, none of them being influential due to the design. For replicated $\mathbf{x}_i$ the reciprocal of the number of replications is the upper bound for h_i, so only for unreplicated treatments can h_i reach the value 1. From equation (4) it is easily seen that if $h_i = 1$ and observation i is removed from the data the design breaks down in the sense that the reduced data does not support the fitting of the intended model. Cook and Weisberg (1982) show that

$$h_i = \frac{1}{n} + \mathfrak{x}'_i(\mathfrak{X}'\mathfrak{X})^{-1}\mathfrak{x}_i, \tag{5}$$

where $\mathfrak{x}_i$ is the i^{th} row of $\mathfrak{X}$, and

$$h_i = \frac{1}{n} + \sum_{l=1}^{p}\frac{(\mathbf{v}_l{}'\mathfrak{x}_i)^2}{\lambda_l} = \frac{1}{n} + \mathfrak{x}'_i\mathfrak{x}_i\sum_{l=1}^{p}\frac{cos^2(\theta_{li})}{\lambda_l},$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ are the eigenvalues of $\mathfrak{X}'\mathfrak{X}$, $\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_p$ are the corresponding eigenvectors and θ_{li} is the angle between $\mathbf{v}_l$ and $\mathfrak{x}_i$. Thus h_i is large if $\mathfrak{x}'_i\mathfrak{x}_i$ is large, that is, $\mathbf{x}_i$ is far away from the bulk of the design points. h_i can also be large when $\mathfrak{x}_i$ is in the direction of an eigenvector corresponding to a small eigenvalue. But note that if $\mathfrak{x}'_i\mathfrak{x}_i$ is small, h_i is small no matter its direction. Thus, for level balanced factors, design points closer to the center of the experimental region have small h_i values as we show in the illustrations in Section 4.

Each h_i can also be expressed as

$$h_i = \mathbf{f}(\mathbf{x}_i)'\mathbf{M}^{-1}\mathbf{f}(\mathbf{x}_i) \tag{6}$$

such that we can write $Var(\hat{y}(\mathbf{x}_i)) = h_i\sigma^2$ where $\hat{y}(\mathbf{x}_i)$ is the estimate of the mean response (estimated from the fitted polynomial) under treatment $\mathbf{x}_i$. This expression establishes the relations between leverage measures and

the G and I criteria (Box and Draper, 1975; Ahmad and Gilmour, 2010) and, together with equation (4), shows the dangers of having high leverage design points. The determinant of the information matrix of the design with design points removed can also be written as a function of the full design information matrix and the elements of the $\mathbf{H}$ matrix, as shown in Andrews and Herzberg (1979), thus giving a link to the D criterion.

For illustration, Table 1 shows a $(DP)_S$ design constructed by Gilmour and Trinca (2012) together with the h_i value for each design point. The design allows 6 degrees of freedom for estimating σ^2 but, if at least one observation from the set of design points 1, 6, 9 and 12 goes missing, the second order model cannot be fitted. Note that when $h_i = 1$ the estimated response from the full data matches exactly its observed value. Even if the situation is not so drastic that the design breaks down, a design point with high leverage may cause inflation in the variances of parameter estimators if its observation goes missing. Thus, when designing an experiment, especially if the number of experimental units is limited, we should caution against the use of design points with high leverage. For p and n fixed Box and Draper (1975) considered minimising the variance of leverages given by

$$n^{-1} \sum_{i=1}^{n} \left(h_i - \frac{p}{n} \right)^2, \tag{7}$$

or its square root, for obtaining designs robust to wild observations. They showed how this measure of robustness varied with the number of center points and the values for the axial points used in CCDs.

Other quantities based on the h_i values are appealing. For example, recall that the contribution of each leverage to Cook's distance is $\frac{h_i}{(1-h_i)^2}$. Thus minimising

$$\frac{1}{n} \sum_{i=1}^{n} \frac{h_i}{(1-h_i)^2} \tag{8}$$

when choosing a design is also a good idea. However, as discussed by Andrews and Herzberg (1979), the use of (7) or (8) or any other measure based only on the leverages as a design criterion may result in very inefficient designs. One explanation for this comes from equation (5) which indicates, as pointed out by Cook and Weisberg (1982), that a design with low h_is benefits from points that are closer to the centre of the region. Since such points have less information for estimating many terms of the polynomial model the design will perform poorly in terms of estimation precision.

Table 1: Designs for Example 1: three 3-level factors in $n = 16$ and $p = 10$

	$(DP)_S$ (2)				H (4)				$\kappa_1 = .5;\ \kappa_5 = .5$ (5)			
i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i
1	−1	−1	−1	1.000	−1	−1	−1	.644	−1	−1	0	.500
2	−1	0	1	.500	−1	−1	0	.613	−1	−1	0	.500
3	−1	0	1	.500	−1	0	−1	.644	−1	0	−1	.729
4	−1	1	−1	.500	−1	0	0	.571	−1	0	1	.729
5	−1	1	−1	.500	−1	0	1	.644	−1	1	−1	.789
6	0	−1	1	1.000	−1	1	0	.613	−1	1	1	.789
7	0	0	−1	.500	−1	1	1	.644	0	−1	−1	.729
8	0	0	−1	.500	0	−1	0	.625	0	−1	1	.729
9	0	1	0	1.000	0	1	0	.625	0	0	0	.500
10	1	−1	−1	.500	1	−1	−1	.644	0	0	0	.500
11	1	−1	−1	.500	1	−1	0	.613	1	−1	−1	.789
12	1	0	0	1.000	1	0	−1	.644	1	−1	1	.789
13	1	1	−1	.500	1	0	0	.571	1	1	−1	.482
14	1	1	−1	.500	1	0	1	.644	1	1	−1	.482
15	1	1	1	.500	1	1	0	.613	1	1	1	.482
16	1	1	1	.500	1	1	1	.644	1	1	1	.482

Following the line of compromising among objectives or design properties, in this paper we consider the construction of compound optimum designs that are efficient in terms of several properties but also control for high leverage. For this last property we choose to use equation (8) because of its relation to common diagnostic techniques and also because we can easily define a design efficiency measure with respect to leverage. Thus equation (8) will be referred to as the H criterion function and the H-efficiency of design X will be calculated by

$$H_{eff} = 100 \left(\frac{\sum_{i=1}^{n} \frac{h_{iH}}{(1-h_{iH})^2}}{\sum_{i=1}^{n} \frac{h_{iX}}{(1-h_{iX})^2}} \right), \tag{9}$$

where the subscript H refers to the H optimum design. Note that a design with at least one point with $h_i = 1$ is 0% efficient under the H criterion. For the theoretical or ideal H-optimum design the numerator in (9) reduces to $\frac{np}{(n-p)^2}$ but such an ideal design rarely exists so we prefer to use equation (9) for measuring the efficiency with respect to the leverages.

Thus, the compound criteria we propose in this paper maximise the compound function

$$\frac{|\mathbf{X'QX}|^{\frac{\kappa_1}{p-1}}(n-d)^{\kappa_4}}{(F_{p-1,d;1-\alpha_1})^{\kappa_1}(F_{1,d;1-\alpha_2})^{\kappa_2}[tr\{\mathbf{W}(\mathbf{X'QX})^{-1}\}]^{\kappa_2+\kappa_3}(\sum_{i=1}^{n} \frac{h_i}{(1-h_i)^2})^{\kappa_5}}, \tag{10}$$

for finding optimal designs, where κ_5 is the weight for H-efficiency and $\sum_{i=1}^{5} \kappa_i = 1$. Choosing different values of the κ_i is subjective and problem dependent, like the choice of utility function in most applications. However, small changes in their values make little difference in practice and small positive weights for particular criteria do not differ much from ignoring these criteria. As suggested by Gilmour and Trinca (2012), it is reasonable to simplify the choice by using relative weights of 3, 1 and 0 for objectives that are considered major, minor or unimportant. When time allows, we would also recommend users to try various weights and to consider all the properties of the designs produced. As with all methods of optimal design, that proposed here should be considered to be mainly a way to produce interesting designs for consideration by the experimental team.

3. An Algorithm

To find the optimum designs we used exchange algorithms. Both point exchange (Cook and Nachtsheim, 1980) and coordinate exchange (Meyer and

Nachtsheim, 1995) versions were explored. The nature of these algorithms is sequentially performing changes in design points or factor coordinates that improve a given initial design, usually a random initial design. The search is performed many times from different initial designs in order to increase the probability of finding the best solution. In this paper we consider candidate design points belonging to the complete factorial design with the minimum number of levels necessary for the response surface model aimed at. A basic description of the algorithm for the coordinate exchange version is given in Algorithm 1 and R (R Core Team, 2016) code is provided in the Supplementary Material.

For criteria that require the calculation of the determinant and/or the inverse of the appropriate information matrix we should use updating formulae after each exchange in step 12. Here we used the methods of Cook and Nachtsheim (1989). If desired, determinants and inverses can be recalculated after each change or after a prespecified number of changes, e.g. to recalibrate calculations in order to avoid building up numerical rounding errors. However, our experience shows that recalibration is unnecessary. The number of degrees of freedom for pure error is n minus the number of distinct treatments. For the point exchange algorithm, a treatment label is attached to each point in the candidate set which is carried forward during the search. For coordinate exchange, treatment labels are attached to the rows of $\mathbf{X}$ after each change in the design.

4. Applications

In this section we show the performances of the proposed criteria in four experimental layouts. We used $\alpha_1 = \alpha_2 = 0.05$ for the $(DP)_S$ and $(AP)_S$ criteria throughout and at least 1,000 tries. The number of tries was increased when the best design appeared just once.

4.1. Example 1

We considered Example 1 from Gilmour and Trinca (2012) and constructed several new designs for the experiment which involved three three-level factors in 16 runs with the second order model as the primary model. In Table 1 we show the $(DP)_S$- and H-optimum designs and a compromise design obtained by composing the $(DP)_S$ and H properties with equal weights. The h_i values for the design points are also shown. As already discussed in Section 2 the $(DP)_S$-optimum design has four points with the upper bound

Algorithm 1: Coordinate Exchange Algorithm

Input: number of factors k; $levels[k][.]$; model terms indicator; design size N; number of tries $Ntries$; weight vector for compound criteria $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_5)$; weight vector for $\mathbf{W}$; significance level for pure error adjustments α_1 and α_2; number of iterations for recalibration $Ncali$

```
for m ← 1 to nTries do
    X ← model matrix from a randomly generated initial design
    crit ← evaluate equation (10)
    crit_0 ← crit
    improve ← 1
    while improve == 1 do
        improve ← 0
        for j ← 1 to k do
            for i ← 1 to N do
                exchange(X[i][j])
                complete all elements of X[i][.] according to the model
                crit_0 ← evaluate equation (10)
                if crit_0 > crit then
                    improve ← 1
                    crit ← crit_0
                else
                    revert_exchange(X[i][j])
    if m == 1 then
        X_best ← X
        crit_best ← crit
    else
        if crit > crit_best then
            X_best ← X
            crit_best ← crit
Return X_best
```

Table 2: Properties of designs for the second order model, Example 1 ($q = 3$, $p = 10$ and $n = 16$)

Design	Criterion	DF (PE;Lof)	Efficiencies					
			D_S	$(DP)_S$	A_S	$(AP)_S$	H	h_{max}
1	D_S; A_S	(0;6)	100.00	0.00	100.00	0.00	50.57	.843
2	$(DP)_S$	(6;0)	83.09	100.00	65.38	83.64	0.00	1.000
3	$(AP)_S$	(5;1)	93.03	96.17	86.27	100.00	26.71	.857
4	H	(0;6)	66.22	0.00	34.85	0.00	100.00	.644
5	$\kappa_1 = \kappa_5 = .5$	(4;2)	92.58	76.13	82.16	81.64	57.40	.789
6	$\kappa_1 = .3; \kappa_5 = .7$	(3;3)	96.62	54.09	93.18	70.471	73.70	.750
7	$\kappa_1 = .5; \kappa_4 = .5$	(5;1)	93.03	96.17	86.17	99.89	26.71	.857
8	$\kappa_1 = .2; \kappa_4 = .8$	(4;2)	95.10	78.21	89.46	88.89	36.68	.832
9	CCD	(1;5)	93.15	1.91	90.75	4.31	43.07	.796

leverage value. For the H-optimum design the best that can be done is choosing all 16 points with leverages ranging from 0.571 to 0.644 (note that the ideal value would be $\frac{10}{16} = .625$). The efficiencies in Table 2 show that this design lacks efficiency in terms of other properties. For the compromise design the h_i values range from 0.482 to 0.789.

These three designs are contrasted with several others in Table 2 (designs 1, 3 and 8 are shown in Gilmour and Trinca (2012), the CCD is the central composite design in a cubic region with two center points). We see that D_S and H single criteria produce designs that do not allow degrees of freedom (DF) for estimating pure error (see designs 1 and 4) while the pure error adjusted criteria produce designs that are poor for lack of fit checking and leverage efficiency (designs 2 and 3). The CCD has just one DF for error and thus has low efficiencies for the adjusted criteria. Designs from 5 to 8 show that we may drive the design by changing the weight pattern in the compound criterion, recalling that κ_1 represents the weight given to inference, κ_4 the weight given to checking for lack of fit and κ_5 the weight given to robustness to missing values. In this case for example we obtained an attractive compromise design (design 5) when considering $(DP)_S$ and H with equal weight, which performs reasonably well in several respects and will not

break down if a point goes missing. Designs 7 and 8 were obtained composing the $(DP)_S$ and DF efficiencies (as in Gilmour and Trinca (2012)) and show that although DF efficiency attenuates leverage problems, the resulting designs are different from those using the H property. Note that design 7 is very similar to design 3 but not equivalent. This highlights the value of constructing several designs and comparing them in terms of a wide range of properties, although in this case the difference is very small.

4.2. Example 2

In this illustration we use as motivation the experiment from Mountzouris et al. (1999) that studied the effects of substrate concentration (X_1), enzyme concentration (X_2) and transmembrane pressure (X_3) on several quantitative characteristics of the product formed (types of oligodextrans). The investigation aimed to fit empirical models, in particular second order polynomials and thus the design used was a three-level CCD with four centre runs ($n = 18$). For one of the treatments, X_1 and X_3 at the highest levels and X_2 at the lowest level (run 15 in Table 3), the reaction did not work and thus the respective run was removed from the data analysis. Fortunately the CCD design with run 15 removed still allowed the fitting of the assumed model. Note however that all points from the two-level factorial have the maximum h_i value that is close to 0.8.

We constructed several alternative designs for this experiment, some of which are shown in Tables 3 and 4. The properties of all designs are shown in Table 5. The ideal value of h_i is $\frac{10}{18} = .556$ and the H-optimum design gives h_i ranging from .514 to .589 but no pure error DF. Despite the high number of distinct treatments in this design, its efficiencies in terms of point estimation are not very high. The other designs constructed by single criteria have very low efficiencies in at least one of the properties studied. Among them, the D_S optimum design has the best performance in terms the H property but only 2 pure error DF. Note that designs 3, 4 and 11 are all $(DP)_S$ optimal but the three designs behave slightly differently with respect to the A_S and $(AP)_S$ criteria, highlighting the importance of evaluating several properties for design choice. Designs from 6 to 13 were obtained by the compound criteria varying the associated weights. The results show that many interesting designs can be constructed and the use of many properties with some weight greater than zero on each is promising as in design 10.

Table 3: Possible designs for Example 2: three 3-level factors in $n = 18$ and $p = 10$

	$(DP)_S$ (3)				H (5)				CCD (14)				$\kappa_1 = \kappa_3 = \kappa_4 = \kappa_5 = .25$ (10)			
i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i
1	−1	−1	−1	.458	−1	−1	0	.564	−1	−1	−1	.794	−1	−1	0	.678
2	−1	−1	−1	.458	−1	−1	1	.589	−1	−1	1	.794	−1	−1	1	.663
3	−1	−1	1	.833	−1	0	−1	.564	−1	0	0	.505	−1	0	−1	.678
4	−1	1	−1	.833	−1	0	1	.514	−1	1	−1	.794	−1	1	−1	.663
5	−1	1	1	.458	−1	1	−1	.589	−1	1	1	.794	−1	1	1	.457
6	−1	1	1	.458	−1	1	0	.514	0	−1	0	.505	−1	1	1	.457
7	0	−1	0	.500	−1	1	1	.589	0	0	−1	.505	0	−1	−1	.661
8	0	−1	0	.500	0	−1	−1	.564	0	0	0	.155	0	0	1	.722
9	0	0	−1	.500	0	−1	1	.514	0	0	0	.155	0	1	0	.722
10	0	0	−1	.500	0	1	−1	.514	0	0	0	.155	1	−1	−1	.663
11	1	−1	−1	.458	0	1	1	.564	0	0	0	.155	1	−1	1	.456
12	1	−1	−1	.458	1	−1	−1	.589	0	0	1	.505	1	−1	1	.456
13	1	−1	1	.833	1	−1	0	.514	0	1	0	.505	1	0	0	.422
14	1	0	0	.500	1	−1	1	.589	1	−1	−1	.794	1	0	0	.422
15	1	0	0	.500	1	0	−1	.514	1	−1	1	.794	1	1	−1	.456
16	1	1	−1	.458	1	0	1	.564	1	0	0	.505	1	1	−1	.456
17	1	1	−1	.458	1	1	−1	.589	1	1	−1	.794	1	1	1	.484
18	1	1	1	.833	1	1	0	.564	1	1	1	.794	1	1	1	.484

Table 4: Possible designs for Example 2: three 3-level factors in $n = 18$ and $p = 10$ (continued)

	D_S (1)				A_S (2)				$(AP)_S$ (4)				$\kappa_1 = .2; \kappa_5 = .8$ (8)			
i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i	X_1	X_2	X_3	h_i
1	−1	−1	−1	.455	−1	−1	−1	.539	−1	−1	−1	.458	−1	−1	−1	.485
2	−1	−1	−1	.455	−1	−1	0	.533	−1	−1	−1	.458	−1	−1	−1	.485
3	−1	−1	1	.708	−1	−1	1	.509	−1	−1	1	.833	−1	−1	1	.662
4	−1	0	0	.506	−1	0	−1	.453	−1	1	−1	.833	−1	0	1	.662
5	−1	0	1	.460	−1	0	1	.506	−1	1	1	.458	−1	1	−1	.485
6	−1	1	−1	.708	−1	1	−1	.715	−1	1	1	.458	−1	1	−1	.485
7	−1	1	0	.460	−1	1	1	.698	0	−1	0	.500	0	−1	0	.616
8	−1	1	1	.523	0	−1	−1	.460	0	−1	0	.500	0	−1	1	.656
9	0	−1	0	.648	0	−1	1	.518	0	0	1	.500	0	0	0	.616
10	0	0	−1	.648	0	0	−1	.487	0	0	1	.500	0	0	1	.600
11	0	1	1	.591	0	1	0	.626	1	−1	−1	.458	0	1	1	.494
12	1	−1	−1	.457	1	−1	−1	.612	1	−1	−1	.458	0	1	1	.494
13	1	−1	−1	.457	1	−1	0	.453	1	−1	1	.833	1	−1	−1	.485
14	1	−1	1	.690	1	−1	1	.616	1	0	0	.500	1	−1	−1	.485
15	1	0	1	.518	1	0	0	.536	1	0	0	.500	1	−1	1	.662
16	1	1	−1	.690	1	1	−1	.455	1	1	−1	.833	1	0	1	.662
17	1	1	0	.518	1	1	−1	.455	1	1	1	.458	1	1	−1	.485
18	1	1	1	.509	1	1	1	.830	1	1	1	.458	1	1	−1	.485

Table 5: Properties of designs for the second order model, Example 2 ($q = 3$, $p = 10$ and $n = 18$)

Design	Criterion	DF (PE;Lof)	Efficiencies					
			D_S	$(DP)_S$	A_S	$(AP)_S$	H	h_{max}
1	D_S	(2;6)	100.00	20.36	99.65	35.14	79.45	.708
2	A_S	(1;7)	99.91	1.64	100.00	4.04	63.79	.830
3	$(DP)_S$	(7;1)	93.14	100.00	84.71	98.91	35.81	.833
4	$(AP)_S$	(7;1)	93.14	100.00	85.65	100.00	35.81	.833
5	H	(0;8)	89.56	0.00	75.63	0.00	100.00	.589
6	$\kappa_1 = \kappa_5 = .5$	(6;2)	79.69	76.75	55.87	60.92	82.00	.724
7	$\kappa_1 = .7; \kappa_5 = .3$	(6;2)	89.46	86.16	74.22	80.92	75.54	.727
8	$\kappa_1 = .2; \kappa_5 = .8$	(5;3)	83.30	68.90	59.06	58.35	87.34	.662
9	$\kappa_1 = .35; \kappa_2 = .35; \kappa_5 = .3$	(6;2)	91.09	87.73	86.03	93.81	67.44	.787
10	$\kappa_1 = \kappa_3 = \kappa_4 = \kappa_5 = .25$	(5;3)	97.98	81.05	94.91	93.77	73.36	.722
11	$\kappa_1 = \kappa_4 = .5$	(7;1)	93.14	100.00	85.07	99.33	35.81	.833
12	$\kappa_1 = .4; \kappa_4 = .6$	(6;2)	94.93	91.43	88.30	96.28	45.70	.807
13	$\kappa_1 = .2; \kappa_4 = .8$	(4;4)	99.69	65.60	97.33	82.43	77.69	.705
14	CCD	(3;5)	84.74	37.96	80.79	52.08	31.75	.794

4.3. Example 3

Subset designs under minimax loss due to missing design points were studied by Ahmad and Gilmour (2010). The loss for design point i was defined as h_i. In their Example 1 they studied several possible subset designs for fitting the four factor second order model in 36 runs. Here we study the properties of their nine designs (Table 5 of Ahmad and Gilmour (2010)) in a cuboidal region and several other alternatives such as D_S, A_S, $(DP)_S$ and $(AP)_S$-optimum designs and some compromise designs obtained by using the compound criterion. Some designs are shown in Tables 6 and 7 and Table 8 shows the properties of the designs.

From the subset designs (designs 1 to 9) the most efficient in terms of leverage is design 4, composed of the S_3 subset plus some center points. It is followed by design 5 and by the modified CCDs which are more efficient in terms of pure error DF. Because the experiment is reasonably large for the model, single criterion optimum designs perform quite well generally except the D_S and A_S optimum designs which result in low pure error DF. Note that the ideal h_i value is 0.417. For the sake of curiosity we also searched for the design that minimizes the maximum h_i value (maximum loss) and found design 16, with no pure error DF. In this example we found design 13, the best in terms of criterion H, only when using a compound criterion. Even trying the single H criterion on 50,000 initial designs, the algorithm returned an inferior design to this one. Again, by using compound criteria we obtained interesting designs for the experiment.

4.4. Example 4

In this example we consider designs with two-level factors, the model including main effects and two-factor interactions. For four factors in $n = 16$ runs, some designs are shown in Table 9. The D_S, A_S and H-optimum designs are the same, the full factorial. Obviously no pure error estimation is possible from this design and thus we cannot estimate σ^2 unbiasedly. The $(DP)_S$ and $(AP)_S$-optimum designs are also equivalent allowing 5 degrees of freedom for pure error. All the six points from this design that are not replicated have $h_i = 1$ and thus the design will break down if at least one of these goes missing. Thus the efficiency in terms of equation (9) is 0%. A compound criterion involving estimation and leverages produced the design with 4 pure error DF that is 24.28% efficient in terms of (9). This could be increased, if desired, by giving more weight to H as in design 4 (see top part of Table 10).

Table 6: Possible designs for Example 3: four 3-level factors in $n = 36$ and $p = 15$

	$D_S; A_S$ (10)					$(DP)_S$ (11)					H (13)				
i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i
1	−1	−1	−1	1	.387	−1	−1	−1	0	.373	−1	−1	−1	−1	.420
2	−1	−1	−1	0	.379	−1	−1	−1	0	.373	−1	−1	−1	−1	.420
3	−1	−1	0	−1	.424	−1	−1	0	1	.397	−1	−1	0	−1	.399
4	−1	−1	0	1	.379	−1	−1	0	1	.397	−1	−1	1	1	.423
5	−1	−1	1	−1	.440	−1	−1	1	−1	.463	−1	−1	1	1	.423
6	−1	−1	1	0	.424	−1	−1	1	−1	.463	−1	0	0	0	.390
7	−1	0	−1	1	.379	−1	0	−1	−1	.373	−1	0	1	1	.401
8	−1	0	−1	−1	.424	−1	0	−1	−1	.373	−1	1	−1	1	.415
9	−1	0	1	1	.424	−1	0	1	1	.401	−1	1	−1	1	.415
10	−1	1	−1	−1	.440	−1	0	1	1	.401	−1	1	0	1	.406
11	−1	1	−1	0	.424	−1	1	−1	1	.450	−1	1	1	−1	.413
12	−1	1	0	1	.424	−1	1	−1	1	.450	−1	1	1	−1	.413
13	−1	1	1	1	.440	−1	1	0	−1	.397	−1	1	1	0	.411
14	−1	1	1	−1	.404	−1	1	0	−1	.397	0	−1	−1	1	.425
15	−1	1	1	−1	.404	−1	1	1	0	.401	0	−1	−1	1	.425
16	0	−1	−1	1	.379	−1	1	1	0	.401	0	−1	0	0	.415
17	0	−1	−1	−1	.424	0	−1	−1	1	.397	0	−1	1	−1	.432
18	0	−1	1	1	.424	0	−1	−1	1	.397	0	−1	1	−1	.432
19	0	0	0	−1	.408	0	0	0	0	.465	0	0	−1	0	.413
20	0	0	1	0	.408	0	0	0	0	.465	0	0	0	1	.427
21	0	1	−1	1	.424	0	1	−1	−1	.397	0	1	−1	−1	.425
22	0	1	0	0	.408	0	1	−1	−1	.397	0	1	−1	−1	.425
23	1	−1	−1	−1	.440	1	−1	−1	−1	.463	0	1	1	1	.419
24	1	−1	−1	0	.424	1	−1	−1	−1	.463	0	1	1	1	.419
25	1	−1	0	1	.424	1	−1	1	−1	.347	1	−1	−1	−1	.413
26	1	−1	1	1	.440	1	−1	1	−1	.347	1	−1	−1	−1	.413
27	1	−1	1	−1	.404	1	−1	1	1	.445	1	−1	1	0	.414
28	1	−1	1	−1	.404	1	−1	1	1	.445	1	−1	1	1	.423
29	1	0	−1	1	.424	1	1	−1	0	.401	1	−1	1	1	.423
30	1	0	0	0	.408	1	1	−1	0	.401	1	0	−1	−1	.416
31	1	1	−1	−1	.404	1	0	−1	1	.401	1	1	−1	0	.413
32	1	1	−1	1	.440	1	0	−1	1	.401	1	1	−1	1	.413
33	1	1	−1	−1	.404	1	1	1	−1	.445	1	1	−1	1	.413
34	1	1	1	−1	.517	1	1	1	−1	.445	1	1	0	−1	.425
35	1	1	1	1	.404	1	1	1	1	.484	1	1	1	−1	.415
36	1	1	1	1	.404	1	1	1	1	.484	1	1	1	−1	.415

Table 7: Possible designs for Example 3: four 3-level factors in $n = 36$ and $p = 15$ (continued)

	$\kappa_1 = .05; \kappa_3 = .10; \kappa_5 = .85$ (15)					H_{max} (16)				
i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i
1	−1	−1	−1	−1	.442	−1	−1	−1	0	.425
2	−1	−1	−1	0	.410	−1	−1	0	−1	.432
3	−1	−1	0	−1	.464	−1	−1	0	1	.432
4	−1	−1	1	1	.428	−1	−1	1	0	.423
5	−1	−1	1	1	.428	−1	0	−1	−1	.425
6	−1	0	−1	−1	.392	−1	0	−1	1	.344
7	−1	0	0	1	.424	−1	0	0	0	.413
8	−1	0	1	0	.440	−1	0	1	−1	.423
9	−1	1	−1	−1	.394	−1	0	1	1	.421
10	−1	1	−1	1	.423	−1	1	−1	0	.344
11	−1	1	−1	1	.423	−1	1	−1	1	.426
12	−1	1	0	0	.392	−1	1	0	−1	.432
13	−1	1	1	−1	.428	−1	1	0	1	.348
14	−1	1	1	−1	.428	−1	1	1	0	.421
15	0	−1	−1	1	.413	0	−1	−1	−1	.430
16	0	−1	−1	1	.413	0	−1	−1	1	.425
17	0	−1	1	−1	.435	0	−1	1	−1	.429
18	0	−1	1	−1	.435	0	−1	1	1	.423
19	0	0	−1	0	.425	0	0	0	1	.413
20	0	1	−1	−1	.415	0	1	−1	−1	.425
21	0	1	0	−1	.435	0	1	−1	1	.344
22	0	1	1	1	.441	0	1	0	0	.413
23	0	1	1	1	.441	0	1	1	−1	.423
24	1	−1	−1	−1	.403	0	1	1	1	.421
25	1	−1	−1	−1	.403	1	−1	−1	0	.430
26	1	−1	0	0	.425	1	−1	0	−1	.431
27	1	−1	1	1	.406	1	−1	0	1	.432
28	1	−1	1	1	.406	1	−1	1	0	.429
29	1	0	−1	−1	.390	1	0	−1	−1	.430
30	1	0	0	1	.445	1	0	−1	1	.425
31	1	1	−1	0	.387	1	0	1	−1	.429
32	1	1	−1	1	.389	1	0	1	1	.423
33	1	1	−1	1	.389	1	1	−1	0	.425
34	1	1	1	−1	.406	1	1	0	−1	.432
35	1	1	1	−1	.406	1	1	0	1	.432
36	1	1	1	0	.373	1	1	1	0	.423

Table 8: Properties of designs for the second order model, Example 3 ($q = 4$, $p = 15$ and $n = 36$)

Design	Criterion	DF (PE,Lof)	Efficiencies D_S	$(DP)_S$	A_S	$(AP)_S$	H	h_{max}
1	$S_4 + 2S_1 + 4S_0$	(11;10)	78.82	70.36	75.03	75.46	71.61	.636
2	$S_4 + S_1 + 12S_0$	(11;10)	67.07	59.88	59.15	59.50	65.09	.658
3	$S_2 + S_1 + 4S_0$	(3;18)	42.29	11.86	30.21	14.53	84.53	.532
4	$S_3 + 4S_0$	(3;18)	89.17	25.02	87.00	41.86	97.93	.438
5	$S_4 + \frac{1}{2}S_3 + 4S_0$	(3;18)	96.78	27.15	92.92	44.70	92.72	.668
6	$S_4 + \frac{1}{2}S_4^{III} + S_1 + 4S_0$	(11;10)	84.95	75.84	70.97	71.39	90.02	.558
7	$S_4 + \frac{1}{2}S_4^{IV} + S_1 + 4S_0$	(11;10)	84.78	75.68	70.65	71.06	89.12	.570
8	$\frac{1}{2}S_4^{III} + S_2 + 4S_0$	(3;18)	71.88	20.17	60.37	29.04	77.53	.669
9	$\frac{1}{2}S_4^{IV} + S_2 + 4S_0$	(3;18)	67.73	19.00	47.72	22.96	87.82	.596
10	D_S; A_S	(4;17)	100.00	41.63	100.00	63.21	99.48	.517
11	$(DP)_S$	(18; 3)	93.67	100.00	86.73	95.74	98.87	.484
12	$(AP)_S$	(16; 5)	95.63	98.51	92.23	100.00	98.08	.520
13	H	(12; 9)	93.39	86.58	87.30	89.60	100.00	.432
14	$\kappa_1 = \kappa_2 = .2;\ \kappa_5 = .6$	(16; 5)	95.25	98.12	90.06	97.65	98.65	.482
15	$\kappa_1 = .05;\ \kappa_3 = .10;$ $\kappa_5 = .85$	(10;11)	95.95	81.89	92.95	91.23	99.43	.464
16	H_{max}	(0;21)	87.69	0.00	83.56	0.00	99.56	.432

Table 9: Possible designs for Example 4: four 2-level factors in $n = 16$ and $p = 11$

	$D_S; A_S; H$ (1)					$(DP)_S; (AP)_S$ (2)					$\kappa_1 = .15; \kappa_5 = .85$ (4)				
i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i
1	−1	−1	−1	−1	.688	−1	−1	−1	−1	0.5	−1	−1	−1	−1	0.690
2	−1	−1	−1	1	.688	−1	−1	−1	−1	0.5	−1	−1	−1	1	0.690
3	−1	−1	1	−1	.688	−1	−1	1	−1	1.0	−1	−1	1	−1	0.833
4	−1	−1	1	1	.688	−1	−1	1	1	0.5	−1	−1	1	1	0.833
5	−1	1	−1	−1	.688	−1	−1	1	1	0.5	−1	1	−1	−1	0.833
6	−1	1	−1	1	.688	−1	1	−1	1	0.5	−1	1	−1	1	0.833
7	−1	1	1	−1	.688	−1	1	−1	1	0.5	−1	1	1	−1	0.690
8	−1	1	1	1	.688	−1	1	1	−1	1.0	−1	1	1	1	0.690
9	1	−1	−1	−1	.688	1	−1	−1	−1	1.0	1	−1	−1	−1	0.464
10	1	−1	−1	1	.688	1	−1	−1	1	0.5	1	−1	−1	−1	0.464
11	1	−1	1	−1	.688	1	−1	−1	1	0.5	1	−1	−1	1	0.690
12	1	−1	1	1	.688	1	−1	1	1	1.0	1	−1	1	1	0.833
13	1	1	−1	−1	.688	1	1	−1	−1	1.0	1	1	−1	1	0.833
14	1	1	−1	1	.688	1	1	1	−1	0.5	1	1	1	−1	0.464
15	1	1	1	−1	.688	1	1	1	−1	0.5	1	1	1	−1	0.464
16	1	1	1	1	.688	1	1	1	1	1.0	1	1	1	1	0.690

Increasing n to 24 runs we obtained the designs shown in Table 11, whose properties are given in the lower part of Table 10. In this case, as n is more than twice the number of parameters in the model, the usual criteria give reasonably efficient designs as does the use of the full factorial plus a half replicate. Note that the H-optimum design allows more degrees of freedom than the $(DP)_S$-optimum design at the cost of some loss of efficiency for estimating the regression parameters. All compound criteria we tried returned a design equivalent either to the $(DP)_S$ or to the H-optimum design, showing that it is good practice to consider several properties in the design criterion.

5. Discussion

Robustness of designs to missing observation has been of concern in the planning of experiments. In this paper we propose the inclusion of a property, the H property, in the compound criteria of Gilmour and Trinca (2012), in order to construct optimum designs that will not break down if an observation goes missing. Similar properties have been used by other authors to evaluate the performances of CCDs or subset designs but have not been used to drive the search of an optimum design. We also confirm that the use of the H

Table 10: Properties of designs for main effects and two-factor interactions, Example 4 ($q = 4$, $p = 11$, $n = 16$ or $n = 24$)

Size	Design	Criterion	DF (PE;Lof)	Efficiencies D_S	$(DP)_S$	A_S	$(AP)_S$	H	h_{max}
$n = 16$	1	D_S; A_S; H	(0; 5)	100.00	0.00	100.00	0.00	100.00	.688
	2	$(DP)_S$; $(AP)_S$	(5; 0)	76.68	100.00	57.32	91.92	0.00	1.000
	3	$\kappa_1 = .2$; $\kappa_3 = .1$; $\kappa_5 = .7$	(4; 1)	81.23	84.09	62.50	93.46	24.28	.875
	4	$\kappa_1 = .15$; $\kappa_5 = .85$	(2; 3)	89.45	28.48	73.30	41.96	49.03	.833
$n = 24$	5	D_S; A_S; $(AP)_S$; $S_4 + \frac{1}{2}S_4$	(8; 5)	100.00	93.90	100.00	100.00	86.19	.583
	6	$(DP)_S$	(11; 2)	90.79	100.00	79.49	87.25	98.85	.557
	7	H	(12; 1)	87.51	99.89	73.50	82.34	100.00	.500

property as single criterion is not interesting because it does not discriminate well between designs. We should highlight that in some examples we were able to find the best H design only when using a compound criterion, even when trying many thousands of initial random designs. This may indicate this criterion alone somehow drives the design to local solutions and perhaps, starting with generally better designs, instead of at random, and improving them in terms of leverages may lead to even better designs.

Although the surrogate measure used in this paper guarantees robustness of the design to just a single missing observation, for reasonably sized designs the approach produces designs robust to three or four missing points. This was shown by a small study carried out for each of the examples and reported in the Supplementary Material. For small n, missing more than one run is likely to lead to very little information being obtained from the experiment. Of course, the failure of a fairly large proportion of the runs to produce a response might in itself tell the experimenters something important about the system under study.

The overall message of this paper is to reiterate the popular advice to consider many properties of factorial and response surface designs before committing to use one for a particular experiment. The compound criterion used here, including the H criterion will allow experimenters to build designs which are robust to missing or outlying observations. We recommend it for

Table 11: Possible designs for Example 4: four 2-level factors in $n = 24$ and $p = 11$

	$D_S; A_S; (AP)_S; S_4 + \frac{1}{2}S_4$ (5)					$(DP)_S$ (6)					H (7)				
i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i	X_1	X_2	X_3	X_4	h_i
1	−1	−1	−1	−1	.396	−1	−1	−1	−1	.557	−1	−1	−1	1	.500
2	−1	−1	−1	1	.583	−1	−1	−1	1	.443	−1	−1	−1	1	.500
3	−1	−1	1	−1	.583	−1	−1	−1	1	.443	−1	−1	1	−1	.500
4	−1	−1	1	1	.396	−1	−1	1	−1	.443	−1	−1	1	−1	.500
5	−1	1	−1	−1	.583	−1	−1	1	−1	.443	−1	1	−1	−1	.438
6	−1	1	−1	1	.396	−1	−1	1	1	.443	−1	1	−1	−1	.438
7	−1	1	1	−1	.396	−1	−1	1	1	.443	−1	1	−1	1	.438
8	−1	1	1	1	.583	−1	1	−1	−1	.473	−1	1	−1	1	.438
9	1	−1	−1	−1	.583	−1	1	−1	−1	.473	−1	1	1	−1	.438
10	1	−1	−1	1	.396	−1	1	1	1	.473	−1	1	1	−1	.438
11	1	−1	1	−1	.396	−1	1	1	1	.473	−1	1	1	1	.438
12	1	−1	1	1	.583	1	−1	−1	−1	.443	−1	1	1	1	.438
13	1	1	−1	−1	.396	1	−1	−1	−1	.443	1	−1	−1	−1	.500
14	1	1	−1	1	.583	1	−1	−1	1	.443	1	−1	−1	−1	.500
15	1	1	1	−1	.583	1	−1	−1	1	.443	1	−1	1	1	.500
16	1	1	1	1	.396	1	−1	1	−1	.443	1	−1	1	1	.500
17	−1	−1	−1	−1	.396	1	−1	1	−1	.443	1	1	−1	−1	.438
18	−1	−1	1	1	.396	1	−1	1	1	.557	1	1	−1	−1	.438
19	−1	1	−1	1	.396	1	1	−1	1	.473	1	1	−1	1	.438
20	−1	1	1	−1	.396	1	1	−1	1	.473	1	1	−1	1	.438
21	1	−1	−1	1	.396	1	1	1	−1	.473	1	1	1	−1	.438
22	1	−1	1	−1	.396	1	1	1	−1	.473	1	1	1	−1	.438
23	1	1	−1	−1	.396	1	1	1	1	.392	1	1	1	1	.438
24	1	1	1	1	.396	1	1	1	1	.392	1	1	1	1	.438

use in practice.

Acknowledgments

The authors are grateful to grant # 2014/01818-0, São Paulo Research Foundation (FAPESP). The first author is grateful to CAPES/MEC/Brazil Government for his scholarship while he was a MSc student in the Biometry Program, Unesp, Botucatu. The authors thank two referees, an associate editor and the editor for helpful comments which led to improvements in the article.

References

Ahmad, T., Akhtar, M., Gilmour, S. G., 2012. Multi-level augmented pairs second order response surface designs and their robustness to missing data. Communications in Statistics - Theory and Methods 41 (3), 437–452.

Ahmad, T., Gilmour, S. G., 2010. Robustness of subset response surface designs to missing observations. Journal of Statistical Planning and Inference 140, 92–103.

Akhtar, M., Prescott, P., 1986. Response surface designs robust to missing observations. Communications in Statistics - Simulation and Computation 15 (2), 345–363.

Andrews, D. F., Herzberg, A. M., 1979. The robustness and optimality of response surface designs. Journal of Statistical Planning and Inference 3, 249–257.

Atkinson, A. C., Donev, A. N., Tobias, R. D., 2007. Optimum Experimental Designs, with SAS. Oxford University Press, Oxford.

Box, G. E. P., Draper, N. R., 1975. Robust designs. Biometrika 62 (1), 347–352.

Box, G. E. P., Draper, N. R., 1987. Empirical Model Building and Response Surfaces. J. Wiley, New York.

Box, G. E. P., Draper, N. R., 2007. Response Surfaces, Mixtures, and Ridge Analyses. J. Wiley, New York.

Cook, R. D., Nachtsheim, C. J., 1980. A comparison of algorithms for constructing exact d-optimal designs. Technometrics 22 (3), 315–324.

Cook, R. D., Nachtsheim, C. J., 1989. Computer-aided blocking of factorial and response-surface designs. Technometrics 31 (3), 339–346.

Cook, R. D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, New York.

Daniel, C., 1976. Applications of Statistics to Industrial Experimentation. J. Wiley, New York.

Ghosh, S., 1982a. Information in an observation in robust designs. Communications in Statistics - Theory and Methods 10 (11), 1173–1184.

Ghosh, S., 1982b. Robustness of designs against the unavailability of data. Sankhyã 44 (B), 50–62.

Ghosh, S., 1983. Influential observations in view of design and inference. Communications in Statistics - Theory and Methods 12 (14), 1675–1683.

Ghosh, S., 1989. On two methods of identifying influential stes of observations. Statistics & Probability Letters 7, 241–245.

Gilmour, S. G., 2006. Response surface designs for experiments in bioprocessing. Biometrics 62 (2), 323–331.

Gilmour, S. G., Trinca, L. A., 2012. Optimum design of experiments for statistical inference (with discussion). Applied Statistics 61 (3), 345–401.

Herzberg, A. M., Andrews, D. F., 1976. Some considerations in the optimal design of experiments in non-optimal situations. Journal of the Royal Statistical Society, Series B 38 (3), 284–289.

Hoaglin, D. C., Welsh, R. E., 1978. The hat matrix in regression and ANOVA. The American Statistician 32 (1), 17–22.

Lu, L., Anderson-Cook, C., Robinson, T. J., 2011. Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier. Technometrics 54 (4), 353–365.

Meyer, R. K., Nachtsheim, C. J., 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. Technometrics 37 (1), 60–69.

Mountzouris, K. C., Gilmour, S. G., Grandison, A. S., Rastall, R. A., 1999. Modeling of oligodextran production in an ultrafiltration stirred-cell membrane reactor. Enzyme and Microbial Technology 24, 75–85.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`

Ridout, M., 2012. Contribution to the discussion of 'Optimum design of experiments for statistical inference (with Discussion)', by S.G. Gilmour and L.A. Trinca. Applied Statistics 61 (3), 376–377.